Exhibit E

Express Mail
Mailing Label No. EV668188067US

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6800–6804, September 1986
Biochemistry

Cloning of a cDNA coding for human factor V, a blood coagulation factor homologous to factor VIII and ceruloplasmin

(cDNA cloning/DNA sequence analysis/blood coagulation)

WILLIAM H. KANE AND EARL W. DAVIE

Department of Biochemistry, University of Washington, Seattle, WA 98195

Contributed by Earl W. Davie, June 12, 1986

ABSTRACT **Coagulation factor V is a high molecular weight plasma glycoprotein that participates as a cofactor in the conversion of prothrombin to thrombin by factor X_a. A phage λgt11 Hep G2 cell cDNA expression library was screened by using an affinity-purified antibody to human factor V, and 11 positive clones were isolated and plaque-purified. The clone containing the largest cDNA insert contained 2970 nucleotides and coded for 938 amino acids, a stop codon, and 155 nucleotides of 3′ noncoding sequence including a poly(A) tail. The coding region includes 651 amino acids from the carboxyl terminus that constitute the light chain of human factor V_a and 287 amino acids that are part of the connecting region of the protein. The predicted amino acid sequence agreed completely with 147 amino acid residues that were identified by Edman degradation of cyanogen bromide peptides isolated from the light chain. During the activation of factor V, several peptide bonds are cleaved by thrombin, giving rise to a heavy chain, a connecting fragment(s), and a light chain. The light chain is generated by the cleavage of an Arg-Ser peptide bond. The amino acid sequence of the light chain is homologous (40%) with the carboxyl-terminal fragment (M_r, 73,000) of human factor VIII. Both fragments have a similar domain structure that includes a single ceruloplasmin-related domain followed by two C domains. The carboxyl terminus of the connecting region, however, shows no significant amino acid sequence homology with factor VIII. It is very acidic and contains a number of potential N-linked glycosylation sites. It also contains about 20 tandem repeats of nine amino acids.**

Human coagulation factor V is a high molecular weight plasma glycoprotein that is required for rapid thrombin formation and normal hemostasis (1). It circulates in blood as a large single polypeptide chain (M_r, 330,000) with little or no coagulant activity (2–6). During the blood coagulation process, factor V is converted to factor V_a by thrombin by limited proteolysis (2–6). This makes available binding sites for factor X_a (7) and prothrombin (8). Factor V_a is composed of a heavy chain (M_r, 110,000) and a light chain (M_r, 76,000), and these two chains are held together by calcium ions (3, 6). The remainder of the original factor V molecule is released as a large connecting fragment(s) that is rich in carbohydrate (3, 6, 9). Factor V_a binds to cell surfaces (9–12) and negatively charged phospholipid surfaces (13–16) through the light chain, and this increases the rate of prothrombin activation ≈10,000-fold by factor X_a (17). Factor V_a is readily inactivated by activated protein C (18), and this results in the cleavage of the heavy chain into two smaller fragments (19). In addition to its role in prothrombin activation, human factor V_a stimulates the activation of protein C by thrombin on phospholipid and cell surfaces (12, 20, 21). Only the light chain is necessary for this reaction (22).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Factor V has a number of physical and biological properties that are similar to factor VIII. For instance, factor VIII is also converted to an activated form (factor $VIII_a$) by thrombin, and this molecule enhances the rate of activation of factor X by factor IX_a. This cofactor effect of factor $VIII_a$ is analogous to that of factor V_a on the activation of prothrombin. Fass *et al.* (23) have also reported amino acid sequence homology between the N-terminal portions of the heavy and light chains of bovine factor V_a and the corresponding peptides from porcine factor $VIII_a$. In addition, these sequences showed amino acid sequence homology with ceruloplasmin, a plasma copper-binding protein (24). Subsequently, both bovine and human factor V have been shown to contain one copper ion per molecule (25). The complete amino acid sequence of human factor VIII has been determined by cDNA cloning (26, 27). It contains a triplicated A domain (≈350 amino acids) with ≈30% internal amino acid homology. These A domains are also ≈30% homologous with the triplicated domains of ceruloplasmin (28). In contrast to ceruloplasmin, factor VIII also contains a connecting peptide (≈900 amino acids) located between the second and third A domains and contains two C domains (≈150 amino acids) located near the carboxyl-terminal end of the molecule.

The biosynthesis of factor V has been demonstrated in human Hep G2 cells (29), bovine aortic endothelial cells (30), and guinea pig megakaryocytes (31). In this manuscript, we report the isolation of a cDNA clone from a Hep G2 cDNA library that codes for the carboxyl-terminal 938 amino acids of human factor V. This includes the entire light chain of factor V_a (651 amino acids) and a portion of the connecting region (287 amino acids).

MATERIALS AND METHODS

Screening of the λgt11 cDNA Library. Human factor V was prepared by the method of Kane and Majerus (4), and rabbit antibodies were prepared and affinity-purified by the method of Canfield and Kisiel (32). The Hep G2 λgt11 cDNA expression library was kindly provided by Frederick S. Hagen (33). The affinity-purified antibody to human factor V was labeled with ^{125}I and then used to screen filter blots as described by Young and Davis (34). Positive clones were also hybridized with a ^{32}P-labeled oligonucleotide probe with a sequence of ACCCAYTCYTGYTCRTACAT, where Y is T or C and R is G or A (35). This oligonucleotide was synthesized by using an Applied Biosystems (Foster City, CA) oligonucleotide synthesizer.

DNA Sequence Analysis. Phage DNA was prepared from positive clones (36, 37). The cDNA inserts were isolated after *Eco*RI digestion. The inserts were then subcloned into pUC9 (38), followed by subcloning of appropriate restriction fragments into M13 phage cloning vectors mp18 and mp19. DNA sequencing was performed by the BAL-31 exonuclease method (39). Dideoxy chain termination sequencing reactions were carried out with ^{35}S-substituted deoxyadenosine 5′-[α-thio]triphosphate (Amersham) and universal M13 prim-

ers as described (39). DNA sequences were stored and analyzed on an Apple Macintosh computer using the DNA INSPECTOR program (Textco, West Lebanon, VT) (40).

Protein Sequence Analysis. The light chain of factor V_a was isolated from 10 mg of human factor V after incubation with the factor V activator from Russell's viper venom (6). The light chain was then reduced, carboxymethylated (41), and digested with cyanogen bromide (42). The cyanogen bromide peptides were purified by gel filtration on Sephadex G-50 followed by HPLC using an Altex Ultrapore C3 reverse-phase separation column as described by McMullen and Fujikawa (41). Seven peptides were isolated in sufficient yield to allow automated amino acid sequence analysis with a Beckman 890C sequencer. Amino acid sequences were analyzed by using the ALIGN computer program of Dayhoff *et al.* (43).

RESULTS AND DISCUSSION

Affinity-purified antibody to human factor V was radiolabeled with ^{125}I and used for the screening of $\approx 2 \times 10^6$ phage from a λgt11 Hep G2 cell cDNA expression library. Eleven positive clones were identified and plaque-purified. The clone containing the largest cDNA insert (λHV2970) also hybridized with ACCCAYTCYTGYTCRTACAT, an oligonucleotide complementary to the DNA that codes for the amino acid sequence of Met-Tyr-Glu-Gln-Glu-Trp-Val. This amino acid sequence is present in the light chain of human factor V_a. The cDNA insert was then analyzed by restriction mapping and sequenced two or more times (Fig. 1). Also, ≈95% of the sequence was determined on both strands. The cDNA insert contained 2970 nucleotides and coded for 938 amino acids, a stop codon (TAG), and 155 nucleotides of 3′ noncoding sequence, including a polyadenylylation signal (AATAAA) and a poly(A) tail (Fig. 2). The predicted amino acid sequence was in complete agreement with 147 amino acids that were identified by Edman degradation of seven cyanogen bromide peptides. These peptides were isolated and purified from the light chain of factor V_a and are overlined in Fig. 2. These data indicate that the cDNA insert in λHV2970 coded for the carboxyl terminus of the protein, including 651 amino acids that constitute the entire light chain of factor V_a (see below) and 287 amino acids that form a portion of the connecting region.

The molecular weight of the light chain of human factor V_a was calculated to be 71,500 without carbohydrate. The addition of the three potential N-linked oligosaccharide chains with a molecular weight of ≈2000 each increases the molecular weight of the glycoprotein to ≈77,500. This is similar to the values of 74,000–78,000, as determined by $NaDodSO_4$/polyacrylamide gel electrophoresis (4–6). The amino acid sequence of the light chain is 40% homologous with the carboxyl-terminal fragment (M_r, 73,000) of human factor VIII (26, 27). Both fragments have a similar domain structure, including a single ceruloplasmin-related A domain followed by two C domains. The A domain in the light chain of factor V_a is ≈40% homologous with the third A domain of both human ceruloplasmin and human factor VIII (Fig. 3A). The A domain of the light chain of factor V_a, however, shows

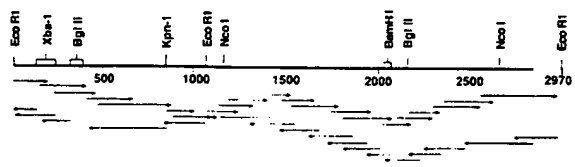

FIG. 1. Partial restriction map and sequencing strategy for the cDNA insert in λHV2970 that codes for human factor V. The extent of sequencing is shown by the length of each arrow, and the direction of the arrow indicates the strand that was sequenced. The coding region is indicated by the solid black line.

less sequence homology (30–36%) with the first and second A domains of factor VIII and ceruloplasmin. Also, the A domain of the light chain of factor V_a contains only a single pair of cysteine residues, whereas the corresponding domains in human factor VIII and ceruloplasmin contain five and three cysteine residues, respectively.

The location and function of the single copper ion associated with factor V remain unknown. Ceruloplasmin contains six copper ions in three types of binding sites (44). Mann *et al.* (25) could not detect type I (blue, 610-nm absorbance) or type III (310-nm absorbance) binding sites in factor V and concluded that the copper may be bound to a type II binding site. By analogy with the type I copper-binding protein plastocyanin, the ligands for the type I copper-binding site at the carboxyl terminus of ceruloplasmin have been proposed to be His-975, Cys-1021, His-1026, and Met-1031 (45). Only two of these four residues are conserved in the light chain of factor V_a (see Fig. 3A). This is in contrast to factor VIII, in which all four residues are conserved in the first and third A domains (26, 27). The carboxyl terminus of ceruloplasmin also contains eight closely clustered histidine residues, which have been suggested to play a role in copper binding (45). Of these eight histidine residues, only three are conserved in factor V (Fig. 2), while five are conserved in factor VIII (26).

The two C domains of human factor V show 35–50% homology with each other and with the C domains of human factor VIII (Fig. 3B). As was noted previously in factor VIII (29), these domains share 20% homology with the slime mold protein discoidin I (46). Discoidin I is a tetrameric galactose-binding lectin, which is essential for cell adhesion in *Dictyostelium discoideum*. It contains the amino acid sequence Arg-Gly-Asp that interacts with a specific cell surface receptor (47). The Arg-Gly-Asp sequence is not present in the two C domains of factor V or factor VIII. The light chain of factor V_a binds to negatively charged phospholipid via electrostatic interactions (14, 16). Pusey and Nelsestuen (16) found that binding was inhibited by 90% when 12% of the lysine residues were modified by citraconic anhydride. It is notable that the C domains in the light chain of factor V_a are basic with a calculated charge at pH 7.0 of +14 (excluding carbohydrate). There are several clusters of basic residues, including a cluster of four or five arginine and lysine residues at the end of each C domain. In contrast, the A domain of the light chain of factor V_a has a calculated charge of −9. A similar charge distribution is seen in the corresponding regions of human factor VIII. It is possible that the binding of the light chain of factor V_a to cell and phospholipid surfaces is mediated by the C domains. However, the precise functions of the various domains in factor V must await future experiments.

During the activation of factor V, the light chain is generated by the cleavage of an Arg-Ser bond located between the connecting region and the light chain. Factor V is rapidly activated by thrombin or the factor V activator in Russell's viper venom. The amino acid sequences determined by Edman degradation of the light chain of factor V_a generated by thrombin (D. B. Wilson, personal communication) or the factor V activator in Russell's viper venom (Fig. 2) are identical, indicating that both enzymes cleave factor V at the same location. The 25 amino acids determined by Edman degradation of the light chain of bovine factor V_a are identical to the sequence of the human protein except for threonine at position 3 and glutamic acid at position 21 (23). Although activation of human factor V with thrombin results in the cleavage of at least three bonds and removal of the large connecting fragment(s), activation with Russell's viper venom results in only one cleavage, which generates the light chain fragment. Factor V activated with Russell's viper venom has a specific activity that is identical to that of factor V activated by thrombin. In contrast, Eaton *et al.* (48) have

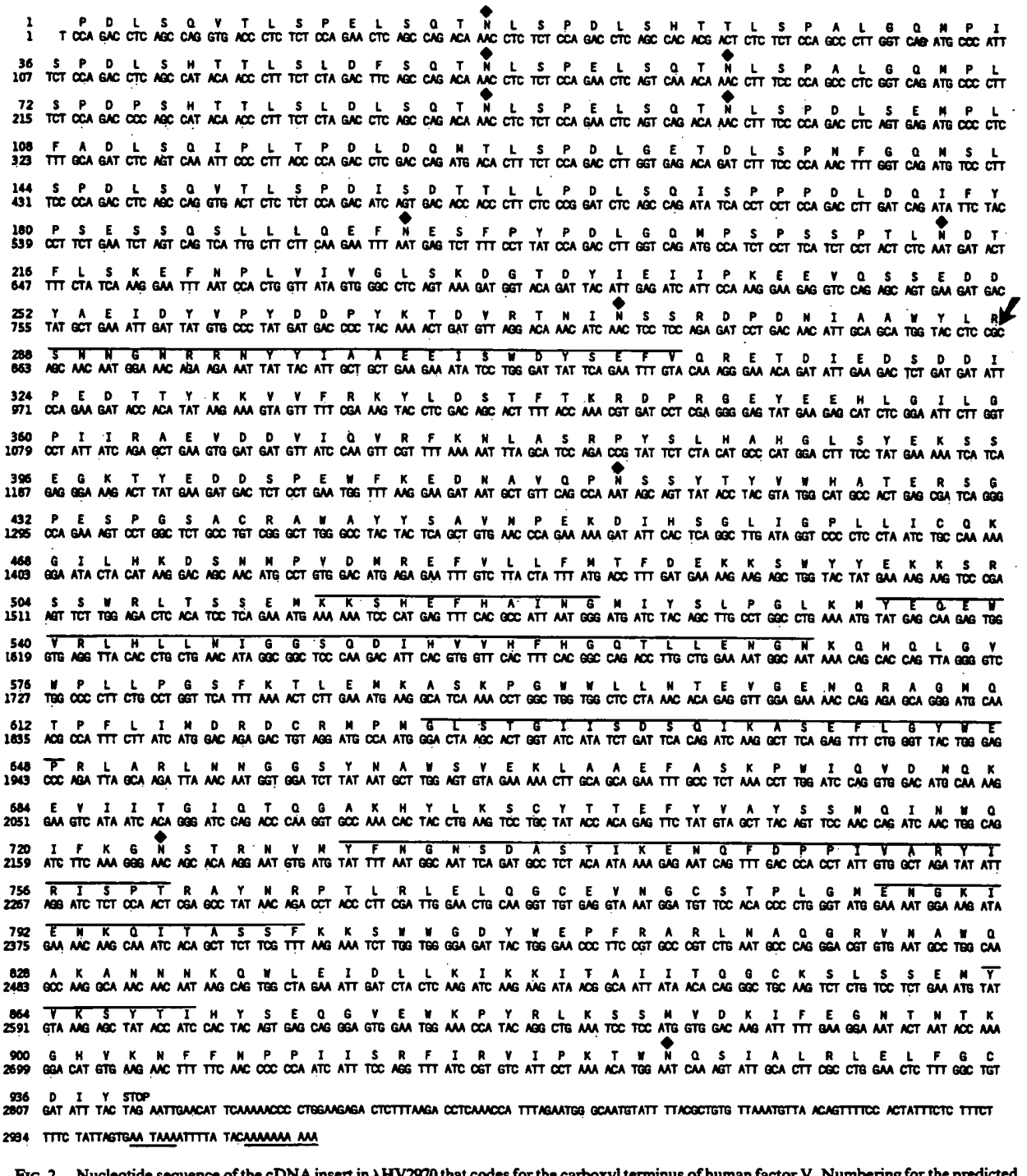

FIG. 2. Nucleotide sequence of the cDNA insert in λHV2970 that codes for the carboxyl terminus of human factor V. Numbering for the predicted amino acid sequence (in single-letter code) is arbitrary and will require revision when the complete amino acid sequence of human factor V is determined. The amino acid residues identified by Edman degradation are overlined. Potential N-linked carbohydrate attachment sites are shown by solid diamonds. A thrombin cleavage site is indicated by a heavy arrow, and the polyadenylylation or processing sequence and the poly(A) tail are underlined.

proposed that cleavage of three peptide bonds is required for complete activation of human factor VIII by thrombin.

The most striking feature of the carboxyl terminus of the connecting fragment of factor V_a is that it contains at least 20 repeats of a sequence of nine amino acids (Fig. 4). The DNA sequence coding for these repeats is strongly conserved and codes for a consensus sequence of Thr-Leu-Ser-Pro-Asp-Leu-Ser-Gln-$\frac{\text{Thr}}{\text{Met}}$. The carboxyl terminus of the connecting fragment (287 residues) is very acidic, with a calculated net charge of −37 at pH 7.0 (excluding carbohydrate). It is also rich in proline (12.2%), serine (15.7%), threonine (8.0%), and leucine (15.2%). It is apparently heavily glycosylated because there are eight potential N-linked oligosaccharide attachment sites within the 287 residues thus far identified. This is consistent with the fact that the connecting fragment(s) does not stain with Coomassie blue and migrates anomalously in

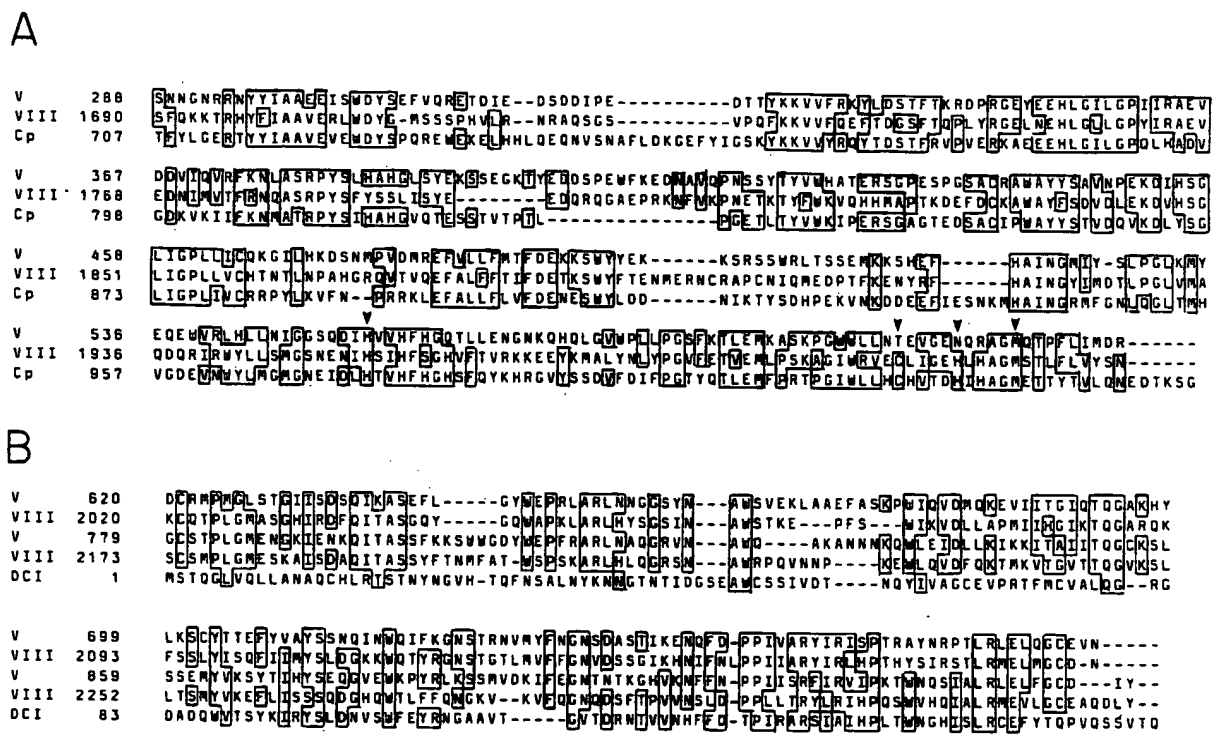

FIG. 3. (*A*) Amino acid sequence (in single-letter code) homology between the A domains in factor V, factor VIII, and ceruloplasmin. Sequences from factor V (residues 288–619), factor VIII (residues 1690–2332), and ceruloplasmin (residues 707–1046) were aligned by using the ALIGN computer program of Dayhoff *et al.* (43). Conserved amino acids are enclosed in boxes. The arrows designate the location of the proposed ligands for type I copper in ceruloplasmin. (*B*) Amino acid sequence homology between the C domains of factor V, factor VIII, and discoidin I. The sequences from factor V (residues 620–778 and 779–938), factor VIII (residues 2020–2172 and 2173–2332), and discoidin I (residues 1–160) were aligned by using the ALIGN computer program of Dayhoff *et al.* (43). Residues conserved in three or more sequences are enclosed in boxes.

$NaDodSO_4$/gel systems (4, 49). The function of this unusual region remains obscure. A search of the Protein Identification Resource* failed to identify any other proteins that are homologous with the 20 repeats of nine amino acids that are present in factor V. Furthermore, the carboxyl terminus of the connecting fragment does not show any significant amino acid sequence homology with the corresponding carbohydrate-rich connecting region of human factor VIII.

In the present investigation, ≈40% of the amino acid sequence of factor V has been established. These findings confirm that factor V, factor VIII, and ceruloplasmin comprise a family of related proteins that have arisen through gene duplication of the A domain (23, 24). Additional cDNA clones must be obtained to determine the domain structure of the remaining 60% of the factor V molecule and the structural features that give rise to its cofactor activity.

*National Biomedical Research Foundation (1986) Protein Sequence Database of the Protein Identification Resource (Natl. Biomed. Res. Found., Washington, DC 20007), Release No. 8.0.

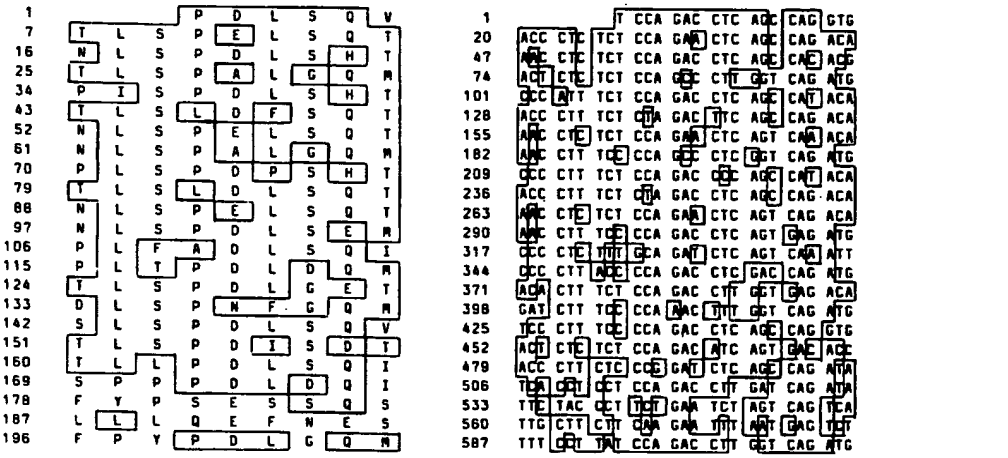

FIG. 4. Tandem repeats in the amino acid sequence (in single-letter code) and the cDNA sequence for the connecting region of human factor V. Amino acid residues or nucleotides that are identical with a consensus sequence of TLSPDLSQT/M or ACCCTTTCTCCAGACCTCAG-TCAGACA are enclosed in boxes.

The authors thank Drs. Donald Foster, Dominic Chung, Kazuo Fujikawa, Akitada Ichinose, Benito Que, Steven Leytus, Barbara Schach, and Kotoku Kurachi for helpful discussions and assistance. We also thank Brad McMullen for assistance in the amino acid sequence analysis, Dr. Fred Hagen for providing the Hep G2 λgt11 cDNA expression library, Dr. Frans Peetoom for kindly making available human plasma for the preparation of factor V, and Mrs. Lois Swenson for her assistance in the preparation of the manuscript. This work was supported by a research grant (HL 16919) from the National Institutes of Health.

1. Owren, P. A. (1947) *Acta Med. Scand. Suppl.* **194**, 1–316.
2. Nesheim, M. E., Myrmel, K. H., Hibbard, L. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 508–517.
3. Esmon, C. T. (1979) *J. Biol. Chem.* **254**, 964–973.
4. Kane, W. H. & Majerus, P. W. (1981) *J. Biol. Chem.* **256**, 1002–1007.
5. Katzmann, J. A., Nesheim, M. E., Hibbard, L. S. & Mann, K. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 162–166.
6. Suzuki, K., Dahlback, B. & Stenflo, J. (1982) *J. Biol. Chem.* **257**, 6556–6564.
7. Freeman, J. P., Guillen, M. S., Bezeaud, A. & Jackson, C. M. (1977) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 675 (abstr.).
8. Esmon, C. T., Owen, W. G., Duiguid, O. L. & Jackson, C. M. (1973) *Biochim. Biophys. Acta* **310**, 289–294.
9. Kane, W. H. & Majerus, P. W. (1982) *J. Biol. Chem.* **257**, 3963–3969.
10. Tracy, P. B., Peterson, J. M., Nesheim, M. E., McDuffie, F. C. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10354–10361.
11. Tracy, P. B. & Mann, K. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2380–2384.
12. Maruyama, I., Salem, H. H. & Majerus, P. W. (1984) *J. Clin. Invest.* **74**, 224–230.
13. Bloom, J. W., Nesheim, M. E. & Mann, K. G. (1979) *Biochemistry* **20**, 4419–4425.
14. van de Waart, P., Bruls, H., Hemker, H. C. & Lindhout, T. (1983) *Biochemistry* **22**, 2427–2432.
15. Higgins, D. V. & Mann, K. G. (1983) *J. Biol. Chem.* **258**, 6503–6508.
16. Pusey, M. L. & Nelsestuen, G. L. (1984) *Biochemistry* **23**, 6202–6210.
17. Nesheim, M. E., Taswell, J. B. & Mann, K. G. (1979) *J. Biol. Chem.* **254**, 10952–10962.
18. Kisiel, W., Canfield, W. M., Ericsson, L. H. & Davie, E. W. (1977) *Biochemistry* **16**, 5824–5831.
19. Walker, F. J., Sexton, P. W. & Esmon, C. T. (1979) *Biochim. Biophys. Acta* **571**, 333–342.
20. Salem, H. H., Broze, G. J., Miletich, J. P. & Majerus, P. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1584–1588.
21. Salem, H. H., Esmon, N. L., Esmon, C. T. & Majerus, P. W. (1984) *J. Clin. Invest.* **73**, 968–972.
22. Salem, H. H., Broze, G. J., Miletich, J. P. & Majerus, P. W. (1983) *J. Biol. Chem.* **258**, 8531–8534.
23. Fass, D. N., Hewick, R. M., Knutson, G. J., Nesheim, M. E. & Mann, K. G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1688–1691.
24. Church, W. R., Jernigan, R. L., Toole, J., Hewick, R. M., Knopf, J., Knutson, G. J., Nesheim, M. E., Mann, K. G. & Fass, D. N. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6934–6937.
25. Mann, K. G., Lawler, C. M., Vehar, G. A. & Church, W. R. (1984) *J. Biol. Chem.* **259**, 12949–12951.
26. Vehar, G. A., Keyt, B., Eaton, D., Rodriguez, H., O'Brien, D. P., Rotblat, F., Oppermann, H., Keck, R., Wood, W. I., Harkings, R. N., Tuddenham, E. G. D., Lawn, R. M. & Capon, D. J. (1984) *Nature (London)* **312**, 337–342.
27. Toole, J. J., Knopf, J. L., Wozney, J. M., Sultzman, L. A., Buecker, J. L., Pittman, D. D., Kaufman, R. J., Brown, E., Shoemaker, C., Orr, E. C., Amphlett, G. W., Foster, B., Coe, M. L., Knutson, G. J., Fass, D. N. & Hewick, R. M. (1984) *Nature (London)* **312**, 342–347.
28. Takahashi, N., Ortel, T. L. & Putnam, F. W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 390–394.
29. Wilson, D. B., Salem, H. H., Mruk, J. S., Maruyama, I. & Majerus, P. W. (1984) *J. Clin. Invest.* **73**, 654–658.
30. Cerveny, T. J., Fass, D. N. & Mann, K. G. (1984) *Blood* **63**, 1467–1474.
31. Chiu, H. C., Schick, P. K. & Colman, R. W. (1985) *J. Clin. Invest.* **75**, 339–346.
32. Canfield, W. M. & Kisiel, W. (1982) *J. Clin. Invest.* **70**, 1260–1272.
33. Hagen, F. S., Gray, C. L., O'Hara, P., Grant, F. J., Saari, G. C., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K. & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2412–2416.
34. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
35. Maxam, A. W. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
36. Silhavy, T. J., Berman, W. L. & Enquist, L. W. (1986) *Experiments with Gene Fusions* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 140–141.
37. Degen, S. J. F., MacGillivray, R. T. A. & Davie, E. W. (1983) *Biochemistry* **22**, 2087–2097.
38. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
39. Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W. & Kurachi, K. (1985) *Biochemistry* **24**, 3736–3750.
40. Gross, R. H. (1986) *Nucleic Acids Res.* **14**, 591–596.
41. McMullen, B. A. & Fujikawa, K. (1985) *J. Biol. Chem.* **260**, 5328–5341.
42. Titani, K., Hermodson, M. A., Fujikawa, K., Ericsson, L. H., Walsh, K. A., Neurath, H. & Davie, E. W. (1972) *Biochemistry* **11**, 4899–4903.
43. Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
44. Ryden, L. & Bjork, I. (1976) *Biochemistry* **15**, 3411–3417.
45. Ryden, L. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6767–6771.
46. Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. (1981) *J. Mol. Biol.* **153**, 273–289.
47. Gabius, H. J., Springer, W. R. & Barondes, S. H. (1985) *Cell* **42**, 449–456.
48. Eaton, D., Rodriguez, H. & Vehar, G. A. (1986) *Biochemistry* **25**, 505–512.
49. Tal, M., Silberstein, A. & Nusser, E. (1985) *J. Biol. Chem.* **260**, 9976–9980.